

A Work Project, presented as part of the requirements for the Award of a Master Degree in Management from the NOVA – School of Business and Economics:



**FROM PUBLICLY AVAILABLE DATA TO BUSINESS DECISIONS  
(ON THE EXAMPLE OF LISBON)**

JONAS SCHWEIGERT 30192

A Project carried out on the Master in Management Program, under the supervision of:

Prof. Miguel Neto

January 2019

## **Abstract**

This study describes the basics of business intelligence, before it explains its standard framework. Subsequently, the paper outlines how Google Maps data can be exploited to create a dashboard in Power BI. Thereby, it applies a modified framework to a case study about venues (businesses, transportation infrastructure, and tourist attractions) in Lisbon, Portugal. Five different analysis are carried out: general KPIs, their correlation, an investigation about popular time as well as heat map and location analysis. This work concludes, that businesses and organizations of all kinds can use this information to identify trends and customer behavior to make data-driven decisions.

## **Keywords**

Business Intelligence | Data | Dashboard | Decision-Making

## Table of Content

Abstract .....	2
Keywords .....	2
1. Introduction .....	4
2. Literature Review .....	4
2.1. Definition and History .....	4
2.2. Data-Driven Decision-Making .....	5
2.3. Process of BI .....	6
2.3.1. Data Sourcing .....	7
2.3.2. Staging .....	8
2.3.3. Data Storage .....	8
2.3.4. Analysis .....	9
3. Methodology .....	10
3.1. Google Maps Data .....	10
3.2. Extraction, Transformation, Load .....	12
3.3. Data Model .....	13
4. Results .....	14
4.1. General Analysis .....	14
4.2. Correlation Analysis .....	15
4.3. Popular Times Analysis .....	17
4.4. Heat Map – Example: Culture .....	18
4.5. Location Analysis – Example: Supermarket .....	20
5. Conclusion .....	22
6. Appendix .....	27
6.1. Appendix A: Python Program to send HTML requests to Google Places API .....	27
6.2. Appendix B: Power BI Dashboard .....	27

## 1. Introduction

The progress of technology and availability of data have led to an increase in data analytics in most industries. As a result, decision-makers are enabled to make data-driven decisions. Although data and tools are available at insignificant low cost, especially small businesses do not take as much advantage as they possibly could. Besides the missing know-how, one of the reasons is certainly the lack of documentation.

The purpose of the study is to draw the line from publicly available data to business decision. The first part of the study defines business intelligence, presents a short overview of its history and outlines the basic BI framework. The framework is divided into several components that are explained generically. The second part consists of a case study about Lisbon. The case study applies the described framework to generate insights into publicly available data about businesses, transportation infrastructure and tourist attractions in the geographical area of Lisbon. The comprehensive data set is retrieved via the Google Places API. Furthermore, the raw data is arranged in a data model which is optimized for analysis. In this study, five different analysis are conducted:

First, the assessment of general key performance indicators (KPI) to benchmark and compare venues. Secondly, a correlation computation of these KPIs with each other. Thirdly, popular time are investigated to detect trends and patterns in user behavior. Fourthly, a heat map with more than 500,000 data points is created to detect crowded geographical areas and lastly a tool is presented that supports decision-making regarding the location of a business on the example of a supermarket.

## 2. Literature Review

### 2.1. Definition and History

Business intelligence refers to the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better

understand its business and market and make timely business decisions (Watson, 2009). Geoff Roberts describes the history of BI in his article “The Rise of Actionable BI” in the following way: BI was first mentioned in 1865 in a Cyclopedia of Commercial and Business Anecdotes. It explains how a banker increased his profits after reacting to his environment prior to his competitors. At that time, data was collected and displayed on paper. Almost one century had passed until IBM developed a new way to store information using magnetic hard disks and an early commercial database. These first computer-aided decision support systems (DSS) were followed by more advanced executive information systems (EIS) and group decision support systems (GDSS), that processed larger amounts of data to enable decision-makers to make information-based decisions. Later, online analytical processing (OLAP) systems as a single location of stored data (data warehouse) were introduced. Massive increases of computation power and rising demands for information have contributed to capabilities to further increase the amount of data. “To solve the problem of time and complexity” in data analysis and visualizations, dashboards were introduced into BI-tools (Roberts, 2012). Due to the development of sensors, mobile devices, and mobile applications, it is now possible even for relatively small organizations to not only collect data about internal operations but exploit data about external circumstances or customer behavior. Consequently, nowadays an increasing number of businesses are exploring new possibilities of uncovering hidden knowledge, improving decision-making, and supporting strategic planning (Roger, 2018).

## **2.2. Data-Driven Decision-Making**

A study by Müller et al. shows, that companies that analyze data increase their productivity by 3 to 7 percent (Müller et al., 2018). This number, however, varies mainly according to the industry and peaks in highly competitive markets. According to Harvard Business Review, data-driven decision-making is concentrated in relatively big organizations with high levels of information technology, an educated workforce, and better awareness for the benefits of data

(McElheran and Brynjolfsson, 2016). However, the progress of technology enables also other companies to extract, store and analyze their data. As the Harvard Review points out, small companies lack the technology on one hand and the know-how for analytics on the other (McElheran & Brynjolfsson, 2016).

It is important for any business, independent of size and industry, to assess key performance indicators considering initial targets to compare the desired performance to the actual outcome. Based on the insights gained, decision-makers can derive improved business actions.

As modern BI systems also enable non-analytics user to analyze and visualize data, decision-making within organizations tends to become increasingly decentralized. Geoff Roberts expects this even to go one step further. He foresees that analysis and insights could be visible immediately due to the evolvement of mobile processing power and BI systems. Thanks to AI, predefined operations can be automated so that there is no additional time needed for human decision-makers to intervene (Roberts, 2012).

### **2.3. Process of BI**

Figure 1 shows the business intelligence framework, that mainly consists of two different components: getting the data in (data warehousing) and taking the data out to generate insights (business intelligence) (Watson and Wixom, 2007).

While the desired analysis and reports are visible in the front-end, the back-end processes are more work-intensive and complex. Before specific context related information can be beneficial to decision-makers, raw data must be extracted, transformed, and loaded into an organized data model to enable analytics.

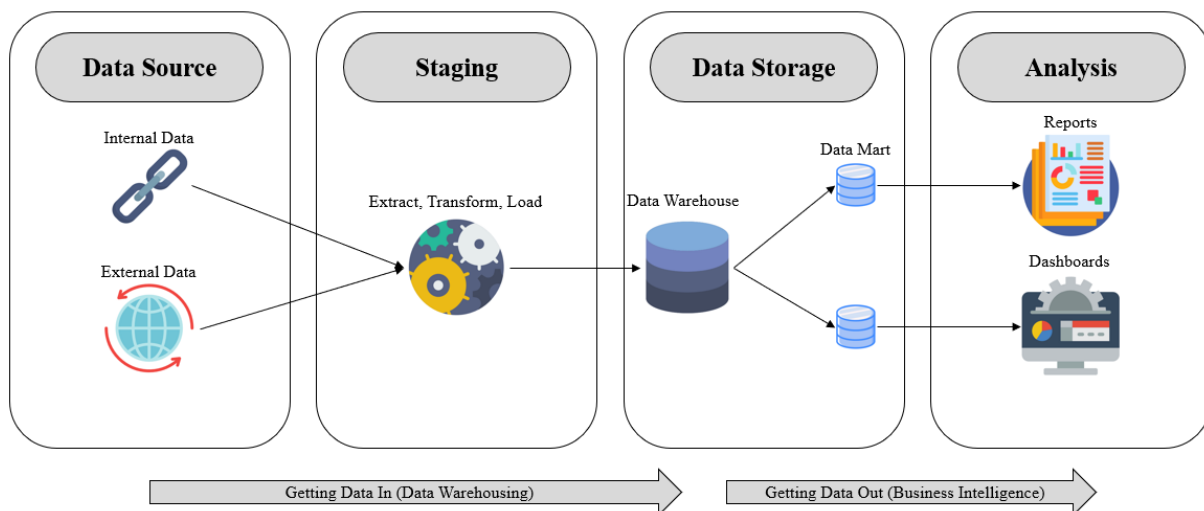


Figure 1: Business Intelligence Framework

### 2.3.1. Data Sourcing

As of today, most companies gather data about their customers, products and services as well as internal operations. This internal data that is gathered within different stages of a company's value chain (E-Procurement, ERP, CRM, SCM). It is mainly used to measure and optimize business processes as well as operational performance. Some companies offer their private data to the public. After signing up, clients of companies like Foursquare or Google can extract data through an API (Application Program Interface) to develop applications or create insights. API's are mostly set up to return pure data responses in the Java Script Object Notation (JSON) format (Blackowl, 2016). The JSON file format is both easy to read for humans and easy to generate, parse and process for machines (json.org, 2018).

Another way of data sourcing is gathering external data obtained organizations through studies, sensors, and simulations and published in online repositories. For instance, the European Data Portal publishes financial and economic data gathered over a timeframe of many years that companies can utilize for analysis at no cost.

Combining external data about the industry with internal data can be used to benchmark activities. Hugh J. Watson points out that it is crucial to assess information not only by itself

but also in its situational context in order to gain a better understanding of the data and identify trends (Watson, 2017).

### **2.3.2. Staging**

In most cases, extracted data from different source systems (extraction) must be modified in the staging area (transformation) before it is loaded into the data warehouse either in batches or as a continuous flow (load). In other words, the goal of this ETL-process is to extract data, that is not optimized for analytics, from different data sources, optimize it for analytics, and eventually move it to a central host (ETL Database, 2018). This process typically consumes a large fraction of the working hours of a data warehousing project (Dayal et al., 2009).

### **2.3.3. Data Storage**

After the modified data is loaded into a data warehouse, this centralized repository “ensures a single version of truth” (Watson, 2007). The storage of the transformed data is subject-oriented, integrated, time-variant, and non-volatile (Inmon, 2005). This implies that within the data warehouse, the data is organized in a specific data model that supports queries and analysis. Furthermore, from now onwards, the data is stored with read-only access and can therefore not be manipulated. Analysis are either run on data queried from a central data warehouse or data marts for a particular functional or geographical area that are being fed with data by the data warehouse.

Data models can be organized in different ways. A star scheme is not normalized and characterized by a high redundancy of data that may lead to bigger file size. A snowflake scheme on the other hand stores data in a normalized form to reduce the duplicated data and with it the file size. The downside of normalization is increasing complexity of the data model in terms of more tables and therefore longer query times. In general, the tables are divided into two different types: fact and dimension tables. Despite storing big amounts of data, fact tables also contain measures. Measures are agile computations that are being calculated while



the analysis is conducted and not physically stored within the file. Hence, it does not have a negative impact on the file size. Dimension tables, on the other hand, contain descriptive information such as types or location that enable analysts to slice the data.

#### **2.3.4. Analysis**

Applications query data from the data warehouse or mart to enable users to create reports and dashboards. A dashboard is “a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance” (Few, 2006). Powerful dashboards are important to facilitate quick and effective decisions. They can offer the right information displayed in an adequate visualization to expedite and enhance the analysis and decision-making process (Firican, 2017). George Firican, a BI specialist from the University of British Columbia, defined key characteristics and best practices that dashboard composers should obey. To enable decision-makers to draw the right conclusions at the right time, the analysis should always include the last refresh date of the underlying data. Moreover, dashboards must display the most important performance indicators and measures. Interactivity, provided through slicers and filters, is a fundamental requirement to engage users and provide further insights. Science has proven that for humans, in general, it is easier to detect trends in visualizations than tables. To enhance the comprehension speed and confidence of users each visualization should contain a title and display information, which is not crucial to draw a conclusion, in tooltips. Due to the reading habits of Westerners, the most important information should always be displayed on the top. Moreover, additional visualizations and chained interactivity should be displayed in a “Z-pattern”. Filters that have an impact on the entire dashboard should be placed on the top ordered by their importance (Firican, 2017). Further investigations of user understanding of visualizations got to the bottom line that simple familiar charts can be interpreted more quickly (Wakeling, 2015).

### 3. Methodology

This study outlines a modified process of BI to generate insights into publicly available data. In the following paragraphs, a case study described, that draws the line from gathering Google Maps data to visualizing insights on a Power BI dashboard on the example of Lisbon. The components of this process are displayed in Figure 2.

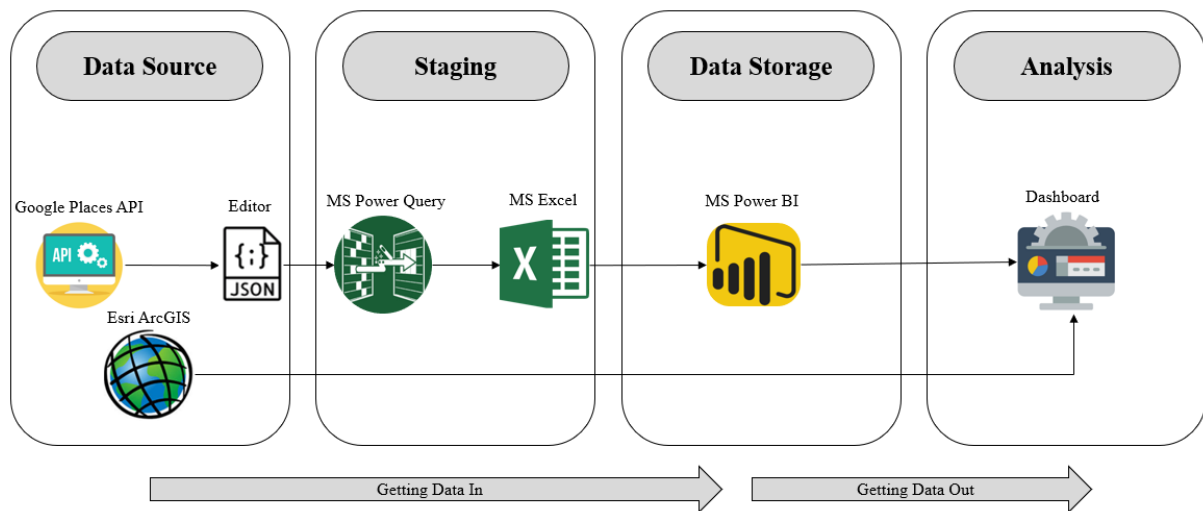


Figure 2: Components of the modified Framework

#### 3.1. Google Maps Data

With regards to private data sets about venues and locations, Google Maps offers the most comprehensive information about all kinds of businesses and public places within a geographical area. Because more than one billion users contribute more than 25 million updates per day it is one of the biggest community-based datasets in the world (Google, 2018). The data about venues is composed of basic information such as the name, address, contact information like the phone number and atmosphere data such as the average rating. However, to sufficiently describe the current state of the behavior and movement of individuals within a geographical area, Google also gathers and displays accumulated data about the popular times (average occupation rate) of each venue as well as the average minimum and maximum time spent by individuals at a venue.

Google offers registered users to access their database through various APIs. Thereby, Google allows every user that is registered with a payment method to access data worth 200\$ for free every month. This is equivalent to data about approximately 2,000 venues. Hence, it is only possible to gather large amounts of data over a period of several months.

The Google Places API offers customers to access basic Google Maps data about venues in a geographical area. As of 2018, the popular times and time spent are not accessible directly through the Google Places API. Thanks to python code, that utilizes a publicly accessible library from GitHub called “populartimes”, it is possible for everyone to also subtract the time-related data. The python program (Appendix A) approaches the API with an HTML request that consists of the API Key (registration-ID assigned by Google), geographical boundaries and a keyword for the type of venues requested. The program output is a JSON file that contains data points further described in table 3.

Name	Assigned by	Type
Google ID	Google	String
Name	Google or modified by venue owner	String
Address	Google or modified by venue owner	String
Google types	Google or modified by venue owner	Strings in Array
Coordinates	Google	Floats in Array
Rating	Google average rating by contributors	Float
Rating number	Number of Google ratings by contributors	Integer
Time spent	Google	Integers in Array
Popular times	Google	Integers in List in Record

Figure 3: Google Places Data

The time that customer spent on average in a venue, as well as the popular times, are based on the average visit duration respectively popularity over the last several weeks (Google, 2018). Thereby, Google uses aggregated and anonymized data from all Android users, that allow Google to access their device’s GPS. According to statscounter GlobalStats Android (Google’s operating system for portable devices) has a market share of more than 72%

worldwide (statscounter GlobalStats, 2018). With approximately 1.85 million inhabitants of the metropolitan area of Lisbon (PorData, 2018) as well as 4.5 tourists (ipdt, 2018) there are annually 4.5 million Android users generating data for the Google Maps database. Thanks to the law of large numbers, first discovered by Jacob Bernoulli, the huge number of users generate a predictable model for user behavior and movement (Bernoulli, 1713).

In addition to Google data, Power BI, a Microsoft tool, offers a plug-in from Esri ArcGIS Maps that supports analysis with contextual demographic data. In contrast to the Google Data, the ArcGIS data is already optimized for analysis and does not require any further modification.

### **3.2. Extraction, Transformation, Load**

The objective of the ETL-process is to retrieve data, optimize it and load it in a data model where users conduct queries and eventually provide insights into the retrieved data. The desired outcome is a correct continuous functionality at adequate performance. Due to the Google Places API limit of requests, the data is only extracted from the source system twice with the last update on the 15<sup>th</sup> November 2018. Nevertheless, the python-routine, as well as the ETL-process, is set up to future data-processing in batches.

The desired raw data must be identified, requested with the geographical area (latitude and longitude) as well as the predefined type of venue, and extracted from Google Places API to make it available for further processing. By sending multiple HTML requests to the Google Places API, the python program extracts the desired records before they are being stored in a text file as a long JSON string. Thus, multiple iterations with different parameters (geographical area and venue type) create several files, that must be unified in a final data file with multiple thousands of records.

Data transformation refers to the preparation of data before the analysis (ETL Database, 2018). Cleansing and aggregation consist of several steps that vary according to the initial condition of the data. Due to the inconsistent modification of the data by various venue owners, this step in the process requires work-intensive standardization, especially for the location data fields. Before the data points can be standardized in Microsoft Excel, the JSON text must be split up into multiple records with the support of Power Query, a Microsoft tool to handle large amounts of data.

The loading of the extracted and transformed data into Power BI is conducted in one batch. Inside Power BI the data model is assembled to enable users to perform the analysis.

### **3.3. Data Model**

The Google Maps database represents an online transaction processing system (OLTP). The main requirement of an OLTP is to collect data from day-to-day transactions. Google's OLTP stores the behavioral data of users anonymously. The transactional data is then transferred and accumulated in an online analytical processing system (OLAP). Through the Google API, it is only possible to retrieve data in an aggregated form from the OLAP. An example of this is the rating of a venue that represents the average of all individual ratings of users (OLAP) instead of a large list of different values – each for a single vote of a user (OLTP).

In this process, the data model is arranged in a mixture of both schemes to reduce complexity in the process on one hand and query times on the other. Due to the split of data, it is organized into several tables that have relationships among themselves.

In Microsoft Power BI analysts can create measures with Data Analysis Expressions (DAX). The DAX functions are designed for user interaction with the reports as they are specially designed to work with relational data and perform dynamic calculations. The exact

computations of the applied measures can be consulted in the Power BI dashboard (Appendix B).

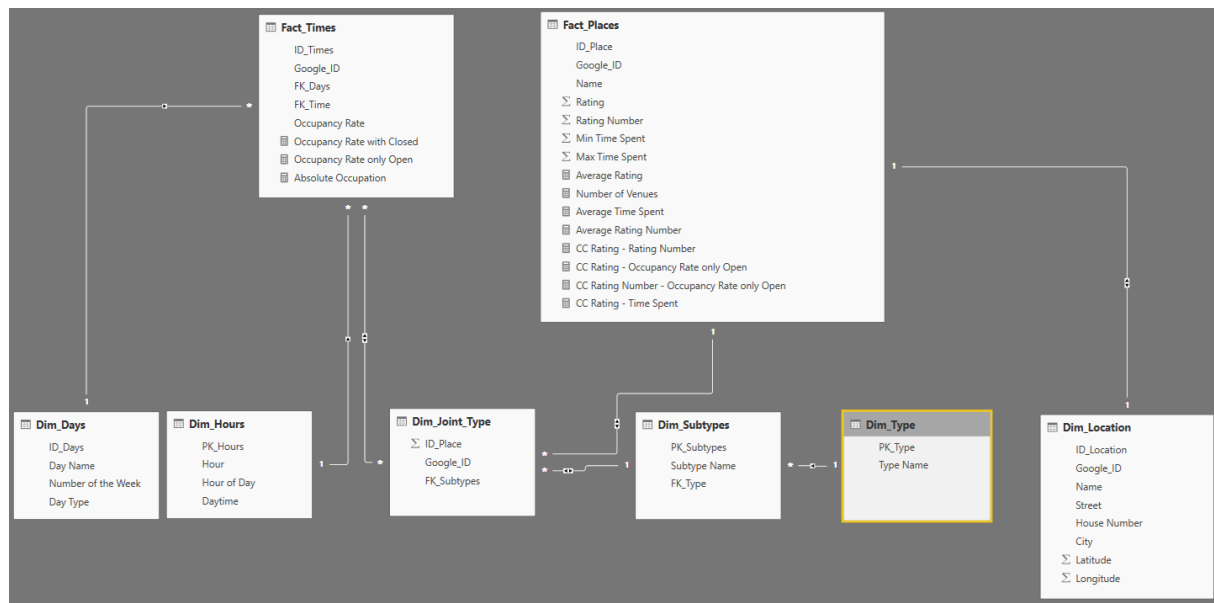


Figure 4: Data Model

## 4. Results

The following paragraph aims to present valuable insights into the constructed data model. Thereby, the analysis follows a predefined order: Idea, Measures, Visualization, Insight, and possible business action. All visualizations presented below can be modified by filters and slicers in the corresponding Power BI dashboard (Appendix B).

The data has last been updated on the 15<sup>th</sup> of November 2018.

### 4.1. General Analysis

*Idea:* At first, it is important to point out the key figures of the given dataset. By displaying the numbers on the top of the first report of the dashboard every user gets a basic understanding to be able to interpret given graphs and compare facts.

*Measures:* Number of Venues, Occupancy Rate only Open, Average Rating, Average Time Spent

*Visualization:* Cards



*Figure 5: KPIs*

*Insight:* In the derived dataset are 3.300 venues that are in the center of Lisbon, Portugal. The Google Maps community rates them with an average rating of 4.17 out of 5 stars. While those locations are open, their average occupancy rate is 37.2 %. On average, customers spent 101 minutes in a venue. However, this number varies drastically according to the type of the venues selected.

*Business action:* The dashboard shows every KPI according to the selected type of business. Thereby, it offers business owners to benchmark their key performance indicators with their competitors' in the geographical area of Lisbon.

## 4.2. Correlation Analysis

*Idea:* The correlation coefficient is used in statistics to measure a relationship between two variables. The value is in the range between -1 and 1 where 1 expresses a strong relationship, 0 no relationship, and -1 a negative relationship. The visualization examines the correlation between the averages of the measures and KPIs of all businesses in the dataset. It is important to keep in mind, that the correlation coefficient does not provide information about the causative relationship between to variables.

*Measures:* Rating, Rating Number, Min Time Spent, Max Time Spent, Average Occupation

Open

*Visualization:* R Correlation Plot

*Insight:* The matrix illustrates extensive statistical computations in an easy to read manner. The blue squares represent a positive relationship between the two examined values. While a missing square is the sign for no statistical relationship, a red one depicts a negative one. The bigger the square, the more related the two evaluated measures.



Figure 6: Correlation Coefficient

Among those KPIs, the Google Maps rating is the most visible characteristic for the audience on Google. The visualization shows a positive correlation between the Google rating and all other observed KPIs. This implies, that venues with a high rating tend to have higher rating numbers (0.1), longer visitor times (0.3), and a higher occupation when the venue is open (0.1). Furthermore, a high rating number has a positive correlation with the minimum time users spent in a venue (0.1) and the occupancy rate when the venue is open (0.2). Considering the correlation between maximum time spent and occupancy rate (-0.1), the statement is true that users prefer to spend more time in a venue when it is less crowded.

*Business action:* Since all metrics seem to be influenced by a positive Google rating, it's a crucial task for business owners of less crowded venues to increase this metric to increase the occupancy rate and therefore the revenue. They could do so by either providing touchpoints such as NFC tags or QR codes for customers to forward them to the venues Google rating



page or giving incentives such as discount or free sample products for submitted Google ratings. According to search engine optimization (SEO) specialists, a higher number and more frequent reviews also result in a higher Google Maps ranking (Heaven Keywords, 2018). Consequently, the increased visibility of a shop or restaurant attracts new potential customers.

### 4.3. Popular Times Analysis

*Idea:* The law of large numbers essentially describes that unpredictable, individual behavior becomes predictable as the number of samples increases massively. Google Maps has gathered billions of data points regarding popular times. Hence, the dataset provides incredible accurate insight into human behavior. Each of the 3.300 venues has 24 data points a day, 7 days a week. Consequently, the dataset includes 554,400 data points that provide information about the occupancy rate of venues in Lisbon.

*Measures:* Occupancy Rate with Closed

*Visualization:* Bar Chart

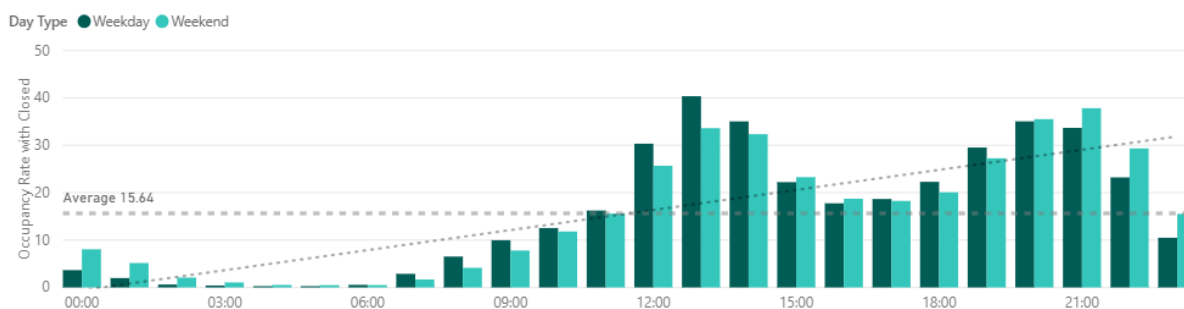


Figure 7: Occupancy Rate per Hour

*Insight:* The bar chart shows the average occupancy rate (Y-axis) of all included venues throughout the day (hours on X-axis). The data is divided into two different bars namely, weekdays (dark) and weekends (bright). This simple visualization facilitates to make a few meaningful statements. Firstly, according to the trendline, throughout the day, venues tend to get more crowded. Secondly, both types of days (weekdays and weekend) peak at 1 PM as

well as at 8 to 9 PM. Thirdly, during the week, venues tend to be more crowded between 6 AM and 8 PM whereas, on the weekend, businesses are more crowded in the evening and night.

*Business action:* Identifying peak and down times of businesses enable decision-makers to allocate staff and resources more efficiently. As the venues in Lisbon tend to be more congested in the afternoons and on weekends, shop owners could rethink and adapt their opening hours.

#### **4.4. Heat Map – Example: Culture**

*Idea:* Heat maps offer a visual summary of geographical datasets that are otherwise difficult to understand for humans. The darker the color, the higher the density of customers in the venues at the time that is being analyzed. This visualization is beneficial for all kinds of businesses that are dependent on high traffic to increase their revenue. This example examines the occupancy rate of venues associated with culture such as museums, art galleries, parks, viewpoints, churches, and other monuments. These places are especially dependent on tourism.

In combination with the bar chart, the heat map becomes even more powerful. While the heat map shows the crowdedness in relation to other venues, the bar chart represents the absolute occupancy of the selected type at each hour of the day. The absolute occupancy is computed by the number of venues of the given type times the occupancy rate. Please note that this measure represents not an exact value, however, it gives a good estimation of the total amount of people at all venues of a given type.

*Measures:* Occupancy Rate only Open

*Visualization:* Custom Visual Heat Map, Bar Chart

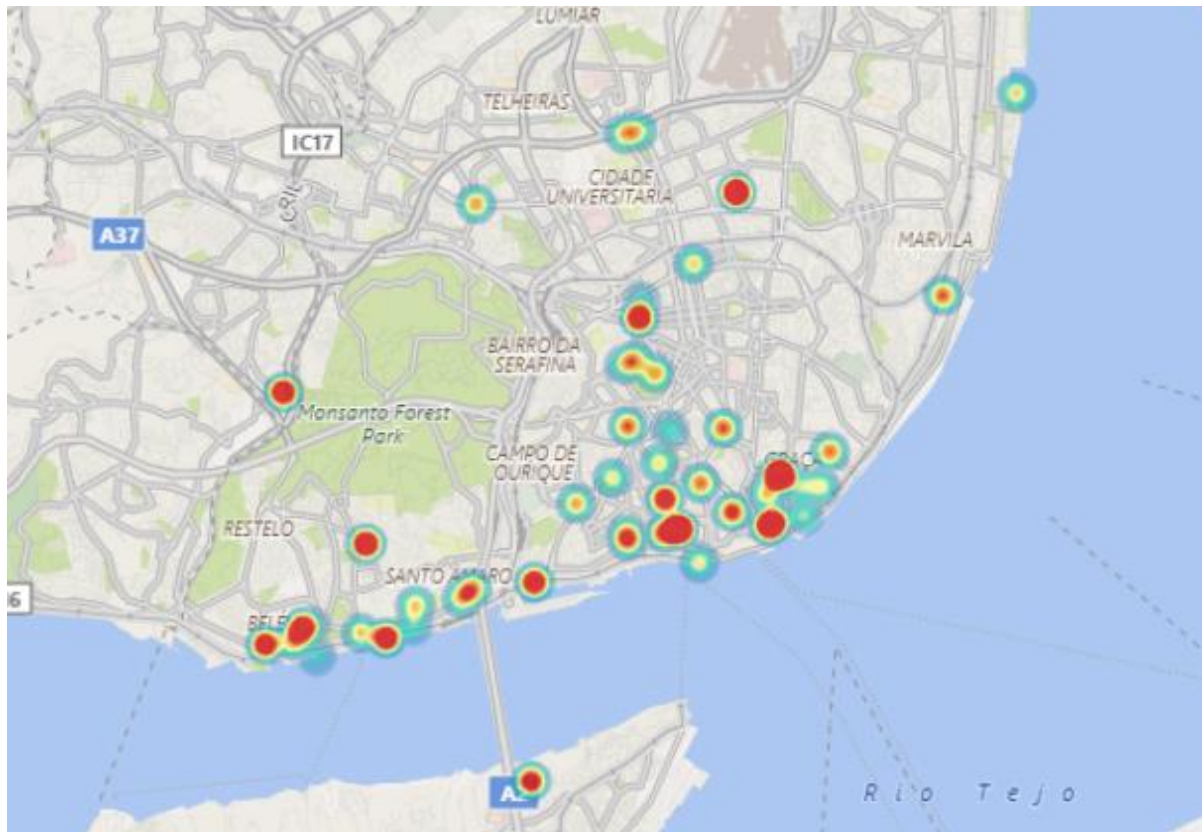


Figure 8: Heat Map

*Insight:* Most venues that belong to the type culture are located in the center of Lisbon and along the river Tejo. As the animation as well as the bar chart illustrate, venues with the type culture generally tend to be crowded from 9 AM to 10 PM. Moreover, it is visible that there is no significant difference in the occupancy rate between weekdays and weekends.

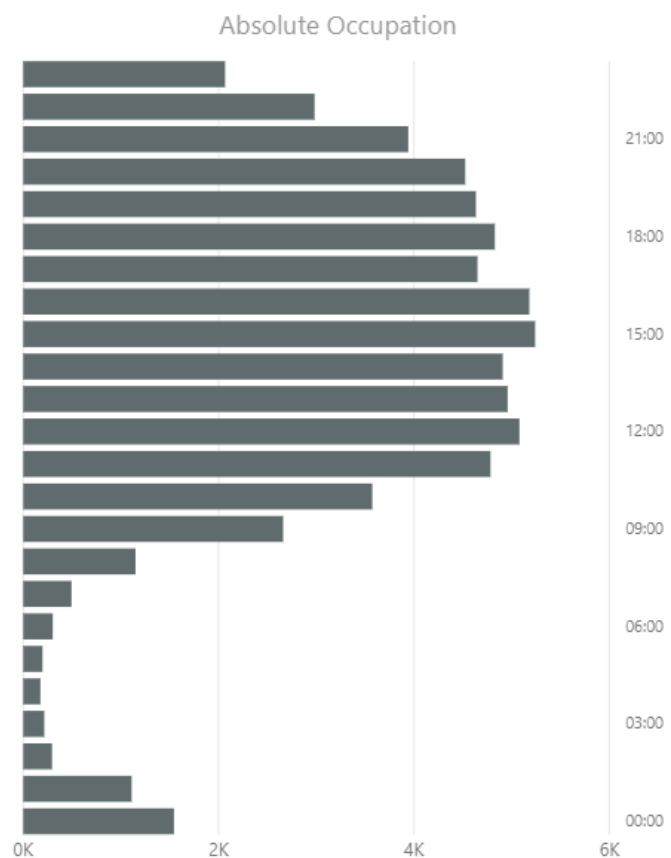


Figure 9: Absolute Occupation

*Business action:* Especially for companies dealing with tourism, the heat map provides valuable insights by illustrating patterns. Businesses that are dependent on revenue generated by tourists may have the possibilities to decided where and when to allocate staff and other resources. In addition to that, they could even consider possible venues to relocate to the crowded areas on the heat map.

In a more general way, the given insights enable marketers to target online ads more precise. Modern social media advertising tools such as Facebook or Instagram offer ads that target users according to their current location. By advertising at the crowded times of a geographical area, the ad reaches out to a maximum number of possible customers. Besides that, heat maps, in a compelling way, can have a positive influence on the decision of investors to allocate financial resources (French, 2014).

#### **4.5. Location Analysis – Example: Supermarket**

*Idea:* Location is crucial for most businesses. This decision implies long-term investment. Relocation of the business is a huge cost. But for some businesses, according to their specific needs, the location is more vital than for others. While retailers and restaurants are dependent on traffic, service businesses mostly perform on client's site. For wholesalers it is important to be close to transportation routes, manufacturing businesses need to be in a good position for transporting supplies and products. For supermarkets, however, multiple factors come into play such as population, demographics, and traffic (Loria, 2017). To use demographic data in a compelling way can have a great influence on the decision to allocate resources. Demographic data, such as the population by gender and age and income, provides valuable insights into the infrastructure needs, resource allocation, and demand for product and services (French, 2014).

Although there are several location analysis tools available in the market, Google Maps data in combination with ArcGIS demographics (published by Michael Bauer Research GmbH) is

a valuable alternative. It provides comparison tools and appealing visualizations at no cost whereas comparable products may offer insights with less preparatory work, but at a much higher cost.

The following guideline, including step-by-step instructions, helps to identify suitable solution among different alternatives to the example of a supermarket location.

*Measures:* Count of Venues, Average Occupancy only Open, Average Rating, Average Rating Number, Average Time Spent

*Visualization:* ArcGIS Map, Tree Map

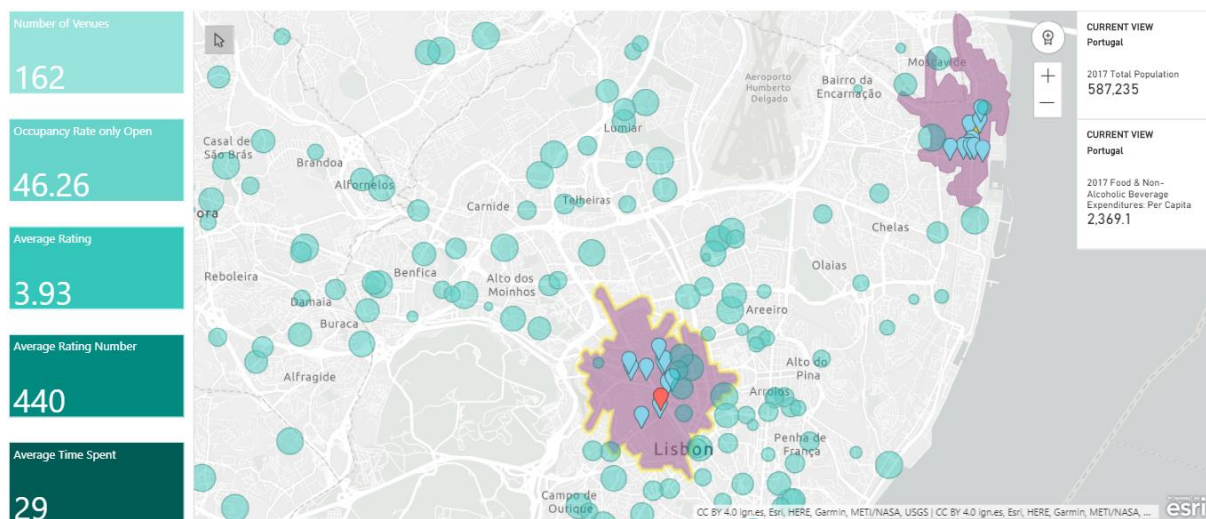


Figure 10: Location Analysis

*Insight:* The ArcGIS Power BI visualization offers a wide range of different features for location analysis. In the beginning, it is crucial to choose the type of business that is being analyzed. As Google distinguishes two types of competing businesses “supermarkets” and “grocery and supermarket”, both must be selected. Next, several possible locations can be inserted by dropping pins on the map. Furthermore, ArcGIS enables the user to find important landmarks such as parking spaces or public transport surrounding the pin (blue pins). A study conducted by Access Development found out that customers travel 8 minutes for grocery shopping (Nanji, 2016). Since most of the customers are walking to the supermarket this can

be translated into a 3-minute drive-time. Selecting the drive-time area shows the number of competitors in the area, their average rating plus the number of ratings as well as the average occupancy when open and the average time spent inside the venue of each customer. These KPIs are displayed in comparison to their overall average in the whole data set of Lisbon. Additionally, infographics present demographic and financial data about the selected area. The table below shows the data presented on the dashboard.

KPI	Lisbon Average	Pin 1: Area “São Sebastião da Pedreira”, Lisbon	Pin 2: Area “Avenida Dom João II, 1990”, Lisbon
Number of venues	162	6	2
Average occupation	46.26 %	44.49 %	46.98 %
Average rating	3.93	3.95	3.95
Average rating number	440	474	225
Average time spent min	29 min	28 min	28 min
Population	852,795	46,687	34,962
Food Expenditures per capita 2017	2,397.62 €	2,768.93 €	2,314.91 €

Figure 11: Comparison of Locations

**Business action:** According to the strategy of the individual business, the decision-makers are better prepared to take an information-based decision according to the insights presented above – or benchmark another location in less than 2 minutes.

## 5. Conclusion

The study aims to enable decision-making with the use of publicly available data. The Google Maps data offers a comprehensive view on businesses, transportation infrastructure and tourist attractions in the city of Lisbon. Before the data can be assessed for possible insights, it needs to be extracted from the Google Places API, cleansed, transformed and loaded into Power BI. Especially the cleansing and transformation of the raw data represent a work-

intense part of the process. Significant visualizations turn the data into information that is easily comprehensible for users. This enables decision-makers to make data-driven instead of gut decisions.

Although there are limitations to the process, decision-makers benefit from the insights in many ways. As of now, there are no comprehensive guidelines, standard processes and documentation available. Additionally, there is neither scientific research about similar processes nor a discussion about advantages and disadvantages of different procedures and tools. Moreover, businesses become dependent on tools suppliers of data and tools if they decide to set up a process to create insights from publicly available data. They may suffer from bugs, missing updates or price increases.

By overcoming those problems and the implementation of a standard process, decision-makers can benefit from data-driven decision-making right away. In this example about the city of Lisbon, decision-makers obtain the opportunity to benchmark their business against competitors, find patterns in customer behavior, and receive decision-support on where to open respectively relocate the business.

To make the most out of the publicly available data, it must be extended by internal company data. The contextual information about the market, customer, competition, and infrastructure plus the internal data gathered in a business' value chain offer customized insights to measure and benchmark the performance. Furthermore, extensive data transformations and insufficient capabilities of geographical visualizations can be either automated or avoided in the recurring cycles. By utilizing low-cost IT-tools and free data, this process offers great possibilities for companies with low financial resources but a bold hands-on mentality.

## References

Bernoulli, Jakob. 1713. *The Art of Conjecturing - Translated into English by Oscar Sheynin*.

Basel: Impensis Thurnisiorum.

Blackowl, Phillip. 2016. *freeCodeCamp.org*. 13. 08. Accessed November 27<sup>th</sup> 2018.

<https://medium.freecodecamp.org/what-is-an-api-in-english-please-b880a3214a82>.

Chiang, Roger. 2018. „Strategic Value of Big Data.“ *Journal of Management Information Systems*, 35: 383-387.

ETL Database. 2018. *ETL Database*. Accessed November 27<sup>th</sup> 2018.

<http://www.etldatabase.com/etl-process/>.

Few, Stephen. 2006. *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly.

Firican, George. 2017. „Best Practices for Powerful Dashboards.“ *Business Intelligence Journal Vol. 22*: 33-39.

French, Dr. Carhlie. 2014. *Why Demographic Data Matters*. Durham: Community Planning New Hampshire .

Google. 2018. *Google*. Accessed November 27<sup>th</sup> 2018.

<https://support.google.com/business/answer/6263531?hl=en>.

Google Cloud. 2018. *Google Cloud*. Accessed November 27<sup>th</sup> 2018.

<https://cloud.google.com/maps-platform/>.

Heaven Keywords. 2018. *How to Rank Higher on Google Maps: 10 Easy Tactics in 2018*.

Accessed November 27<sup>th</sup> 2018. <https://medium.com/@KeywordsHeaven/how-to-rank-higher-on-google-maps-683fa645c414>.



- Hugh J. Watson, Barbara H. Wixom. 2007. „The Current State of Business Intelligence.“ *IEEE Computer Society*: 96-99.
- Inmon, William H. 2005. *Building the Data Warehouse 4th Version*. Wiley.
- ipdt. 2018. *Tourism Report*. Lisbon: Institute for Tourism Planning and Development Portugal.
- json.org. 2018. *json.org*. Accessed November 27<sup>th</sup> 2018. <https://www.json.org/>.
- Loria, Keith. 2017. *Grocery Dive*. Accessed November 27<sup>th</sup> 2018. <https://www.grocerydive.com/news/grocery--how-do-grocery-stores-find-the-right-location-for-expansion/535437/>.
- McElheran, Kristina, and Erik Brynjolfsson. 2016. *HBR.org*. 03. 02. [https://hbr.org/2016/02/the-rise-of-data-driven-decision-making-is-real-but-uneven?referral=03759&cm\\_vc=rr\\_item\\_page.bottom](https://hbr.org/2016/02/the-rise-of-data-driven-decision-making-is-real-but-uneven?referral=03759&cm_vc=rr_item_page.bottom).
- Müller, Oliver, Jan vom Brocke, und Maria Fay. 2018. „The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics.“ *Journal of Management Information Systems*, 35: 1-41.
- Nanji, Ayaz. 2016. *How Far Consumers Travel to Make Everyday Purchases*. Accessed November 27<sup>th</sup> 2018. <http://www.marketingprofs.com/charts/2016/31051/how-far-consumers-travel-to-make-everyday-purchases>.
- PorData. 2018. *Resident Population*. Retrieved from PorData database. <https://www.pordata.pt/en/Municipalities/Resident+population++according+to+the+Census+total+and+by+size+of+population+centre-24>.
- Roberts, Geoff. 2012. „The Rise of Actionable BI: From Paper Binders to Mobile Dashboards and Beyond.“ *Business Intelligence Journal*: 15-19.

- Simon Wakeling, Paul Clough, James Wyper, Amy Balmain. 2015. „Graph Literacy and Business Intelligence: Investigating User Understanding of Dashboard Data Visualizations.“ *Business Intelligence Journal* Vol. 20: 8-19.
2018. statscounter GlobalStats. 11. <http://gs.statcounter.com/os-market-share/mobile/worldwide>.
- Umeshwar Dayal, Malu Castellanos, Alkis Simitsis, Kevin Wilkinson. 2009. „Data Integration Flows for Business Intelligence.“ *EDBT '09*. Saint Petersburg, Russia: ACM: 1-11.
- Watson, Hugh J. 2017. „Adding Context to Information.“ *Business Intelligence Journal*: 5-9.
- Watson, Hugh J. 2017. „Data Visualization, Data Interpreters, and Storytelling.“ *Business Intelligence Journal* Vol. 22: 5-10.
- Watson, Hugh J. 2009. „Tutorial: Business Intelligence – Past, Present, and Future.“ *Communications of the Association for Information Systems*: 487-511.
- Watson, Hugh J., and Barbara H. Wixom. 2007. „The Current State of Business Intelligence.“ *IT Systems Perspective*: 96-99.

## 6. Appendix

### 6.1. Appendix A: Python Program to send HTML requests to Google Places API

```
import populartimes
import json
import time
import random

#allowed place type: https://developers.google.com/places/supported\_types

def call (type_name):
    request = populartimes.get("*Your Google API Key*",
                               [type_name],
                               (38.688101,-9.256214),
                               (38.784693,-9.091105))
    return request
"""
types = ["airport"d,"amusement_park"d,"aquarium"d,"art_gallery"d,"atm"d,"bakery","bank","bar"d,"beauty_salon",
        "bicycle_store",
        "book_store","bus_station"d,"cafe"d,"campground","car_rental"d,"casino"d,"cemetery"d,"church"d,"city_hall",
        "clothing_store","convenience_store","dentist","department_store","doctor",
        "electronics_store","embassy","fire_station","furniture_store","gas_station","gym"d,"hair_care",
        "hardware_store","home_goods_store","hospital","insurance_agency","jewelry_store","laundry"d,"library",
        "liquor_store","lodging","meal_delivery","meal_takeaway"d,"mosque","movie_theater","museum"d,
        "night_club"d,"park","parking","pet_store","pharmacy","police","post_office","real_estate_agency",
        "restaurant"d,"rv_park","school","shoe_store","shopping_mall","spa"d,"stadium","storage"d,"store",
        "subway_station"d,"supermarket"d,"taxi_stand"d,"train_station"d,"transit_station","travel_agency","zoo"d]
"""
types = ["convenience_store"]

for typ in types:
    print("Start with: " + typ)
    type_text = call(typ)
    typ_json = json.dumps(type_text)
    #C:\Users\jonas\Desktop\Files
    with open("C:/Users/jonas/Desktop/Files/" + typ + ".txt","w") as outfile:
        json.dump(typ_json,outfile)

    print("Done with: " + typ)
    time.sleep(random.randint(10,16))
```

### 6.2. Appendix B: Power BI Dashboard

#### Possibility 1:

The Power BI dashboard (as a “.pbix” file) plus the corresponding Excel files can be downloaded in the following Google Drive repository:

[https://drive.google.com/open?id=1tRUny4nDyQUFaGTu26\\_m01vFY7OtDo9V](https://drive.google.com/open?id=1tRUny4nDyQUFaGTu26_m01vFY7OtDo9V)

Please note: To facilitate the capabilities of Esri ArcGIS Maps, users have to sign up for a 60 days free trial.

**Possibility 2:**

1. Go to <https://powerbi.microsoft.com/en-us/>.
2. Press “Sign in” in the top right corner with your NOVA SBE credentials.
  - a. For students: Please add “students” to your email address ([12345@students.novasbe.pt](mailto:12345@students.novasbe.pt)).
  - b. For other roles: Please add your role to your email address
3. Sign up for the “60 Days Pro Trial”.
4. Scan QR-code or open report link in your browser:
  - a. QR code either with QR scanner or Power BI mobile app:



- b. Link for web-browser.

<https://app.powerbi.com/groups/me/reports/ae3b2779-0fdf-424a-92af-a866107036ae?ctid=c42931e3-6733-41e6-8d42-c48ba334bba4>